



Borra, Virginia Laura
Marfetán, Diego
Mignoni, César Antonio
Pagura, José Alberto
Rampello, Yamila

Instituto de Investigaciones Teóricas y Aplicadas, Escuela de Estadística

COMPARACION DE ESTRATEGIAS DE MUESTREO EN LA ESTIMACION DEL AREA CUBIERTA POR BOSQUE NATIVO¹

En la estimación de valores poblacionales a partir de muestras de poblaciones finitas, la información auxiliar puede constituir un aporte de importancia en la mejora de los planes de muestreo, ya sea incorporándola en la fase de estimación o aprovechándola en el procedimiento de selección de las unidades con el recurso del enfoque basado en modelos o enfoque de predicción. En este marco teórico, la población finita se considera una muestra de una población infinita o superpoblación y se plantea un modelo teórico que refleja el comportamiento de la variable de interés.

Cuando las unidades de muestreo se encuentran ubicadas en el espacio puede ocurrir que las cercanas presenten valores similares de la variable en estudio, y tengan importantes diferencias con las que se encuentran a mayor distancia. Este comportamiento es conocido como correlación espacial y puede modelarse mediante correlogramas o semivariogramas, los que pueden tenerse en cuenta para definir el estimador de acuerdo al enfoque de predicción o para decidir un procedimiento eficiente de selección de la muestra.

En este trabajo se estudia el uso de la información proporcionada por los modelos de correlogramas en la selección de la muestra, cuando interesa la estimación de un total o un promedio empleando el método de simple expansión. El problema que se aborda es el de la selección de la muestra para la estimación del área cubierta por bosque nativo en una región de 90Kms.x90Kms. que abarca tres departamentos de la República Oriental del Uruguay. La superficie se ha dividido en unidades de muestreo de 1km.x1km., quedando conformada la población por 8100 unidades. La variable "Y" que se observa a cada una de ellas, es la superficie cubierta por bosque nativo dada por una imagen de satélite.

De esta manera, las unidades de la población quedan organizadas en M=90 filas y N=90 columnas y el estimador del total poblacional será $\hat{Y} = MN\hat{\bar{Y}}$, donde $\hat{\bar{Y}}$ es el estimador de la media \bar{Y} de la población finita.

Los métodos de selección usuales para una población como la que se considera y aplicados en este trabajo, son: muestreo aleatorio simple, muestreo estratificado, conformando los estratos mediante bloques de unidades contiguas y muestreo sistemático con varios arranques aleatorios. Las comparaciones entre los métodos se realizan considerando el enfoque basado

¹ Trabajo elaborado en el marco del Proyecto UNR 80020180300081UR titulado "Planes de muestreo para poblaciones con variabilidad espacial. Propuestas y estudio de propiedades", dirigido por Dr. José Alberto Pagura



en modelos y mediante el error cuadrático medio del estimador:
 $ECM(\hat{Y}) = E\left[V\left(\hat{Y}\right)\right] + E\left[\bar{Y} - \mu\right]^2$, donde μ es el valor medio de la variable en la superpoblación.

El segundo término no depende de la localización de las unidades incluidas en la muestra, mientras que el primero es función del correlograma, cuyos valores dependen de la distancia entre las unidades de la muestra. Como consecuencia, el tamaño del bloque y el período en la selección sistemática estarán proporcionando diferentes errores cuadráticos medios. De esta forma, se considera que el plan más eficiente será aquel en el que $E\left[V\left(\hat{Y}\right)\right]$ sea menor.

Dado que la población se encuentra organizada en 90x90 retículas se consideran bloques cuadrados de lado 9, 18 y 45. Por ejemplo si el tamaño de bloque es 9x9 habrá 100 bloques o estratos. El tamaño de muestra considerado es 200: en el muestreo aleatorio simple se eligen 200 unidades al azar, en muestreo estratificado se emplea la adjudicación proporcional y en muestreo sistemático se utilizan tantos arranques aleatorios como el tamaño muestral dividido la cantidad de bloques.

Se utilizan las esperanzas de las variancias del estimador en términos del correlograma y se calculan las eficiencias relativas de los diferentes métodos. El muestreo sistemático resulta más eficiente que el muestreo estratificado y el muestreo aleatorio simple, aunque esa diferencia se vuelve poco relevante al utilizar pocos estratos y una mayor cantidad de arranques aleatorios para las muestras sistemáticas.

REFERENCIAS BIBLIOGRÁFICAS

- Ambrosio Flores, L. (2000) *Estadística Espacial*. Servicio de Publicaciones de la Escuela Técnica Superior de Ingenieros Agrónomos, Madrid.
- Ambrosio Flores, L. (2015) *Muestreo*. Servicio de Publicaciones de la Escuela Técnica Superior de Ingenieros Agrónomos, Madrid.
- Benedetti, R.; Piersimoni, F.; Postiglione, P. (2015) *Sampling Spatial Units for Agricultural Surveys*. Springer.
- Cressie, N. (1993). *Statistics for Spatial Data*. John Wiley & Sons.
- R Core Team (2019). R: A Language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.